DOCUMENT RESUME

ED 102 974 IR 001 656

AUTHOR Harline, N. La Var
TITLE A Computerized Reference Retrieval System.
INSTITUTION Brigham Young Univ., Provo, Utah. Graduate Dept. of Library and Information Sciences.
PUB DATE Aug 74
NOTE 59p.

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS *Biographical Inventories; *Data Bases; *Information Retrieval; Information Systems; Library Automation; *Library Reference Services; Programing; *Reference Materials; Relevance (Information Retrieval); Research Tools

ABSTRACT

An automated retrieval system for reference works was constructed as a prototype in order to illustrate the capabilities and potentialities of future retrieval systems. More than 170 biographical works were indexed and stored in a computer disk to form a biographical data base, and a computer program was written to match patron queries. Consistent with the indexing, the system retrieved appropriate reference sources to biographical reference problems. (Author/SK)

BEST COPY AVAILABLE

ED102974

A COMPUTERIZED REFERENCE RETRIEVAL SYSTEM

BEST COPY AVAILABLE

A Research Paper

Submitted to The

Graduate Department of Library and Information Sciences

Brigham Young University

Provo, Utah

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

In Partial Fulfillment

of the Requirements of

L.I.S. 697

by
N. La Var Harline
August 1974

IR 001 656

BEST COPY AVAILABLE

ABSTRACT

The problem was to construct an automated retrieval system for biographical reference works. This retrieval system, although limited in scope was to serve as a prototype of future retrieval systems and to illustrate the capabilities and potentialities of computerized reference retrieval systems. Over 170 biographical reference works were indexed in terms of three categories; namely, (1) miscellaneous, (2) occupation, and (3) time period. These works were then stored on a computer disk; thus establishing a biographical data base. A computer program was then written to match patron queries with this data base. Consistent with the indexing, the system retrieved appropriate reference sources to biographical reference problems.

ACKNOWLEDGMENTS

The author wishes to express his deep appreciation to all who have shown an interest in this project. Special thanks go to Professor Keith Stirling who has so ably and unselfishly devoted his time and talents in helping the author to bring this project to fruition. The author also gives thanks to Mr. Bruce Smith, Mr. Don Novak and Mr. Larry Gibson, all of whom are students or graduates of the Computer Science Department of BYU, and who have directed the author in the writing of the computer program. Also a very special appreciation goes to my lovely wife who has been so patient, understanding, and encouraging during the course of this project.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

TABLE OF CONTENTS iii

LIST OF FIGURES iv

CHAPTERS

I. THE PROBLEM 1

Introduction
Statement of the Problem
Delimitation of the Problem

II. LITERATURE SEARCH 4

Examination of Weil's Retrieval System
Examination of Meredith's Retrieval System
Summary and Conclusions of Literature Search

III. BUILDING THE RETRIEVAL SYSTEM 13

Methodology Statement
Devising a Classification Scheme
Authority List Established
Indexing the Works
Establishing the Data Base on Disk
Writing the Computer Program
Variations of Search Routines

IV. SUMMARY AND CONCLUSIONS 32

Recommendations

APPENDIX A 36

APPENDIX B 40

APPENDIX C 48

SOURCES CONSULTED 53

LIST OF FIGURES

Chart | Page

1. Descriptor Classification Scheme 18-20

2. Search Examples 49-52

CHAPTER I

THE PROBLEM

To most librarians it is very evident there is a tremendous volume and variety of reference sources available in today's modern research libraries. Bibliographic control of these reference works is difficult to maintain in a manner simple and understandable to the novice library user. Because of this, patrons are constantly confronted by the problem of which reference book to search to find information about various reference problems. Hence, the patron generally stumbles around on his own, or asks a reference librarian. The librarian, in turn, attempts to recall, either by memory, personal notes, or reference books, which type of reference works are most appropriate to the patron's given problem. Sometimes, despite his training and experience, even the librarian is unable to suggest accurately and speedily all the possible type of sources the patron may search to find an acceptable answer.

The tremendous speed and memory capabilities of a computerized retrieval system can render assistance in suggesting various types of tools the patron may want. Ideally, all the reference books (of a given library, for example) could be indexed and stored on a rapid access device so that a patron's question could then be fed into the system. Quickly and accurately, the computer would print out all the possible reference sources from which a patron could find an acceptable

answer. This project discussed such an automated system.

Statement of the Problem

Briefly stated, the problem was to construct an automated retrieval system for biographical reference works. The paper was partitioned into two different parts. The first part of the paper, the literature search, concerned itself with the following question: What is involved in implementing an automated reference system? In other words, generally speaking, what problems occur in the implementation of any computerized reference system?

The second part of this paper concerned itself with the construction of an automated reference search system based upon the elements studied in the first part of the paper. This system was not all things to all people. On the contrary, the system was extremely limited, both in purpose and scope. The main purpose of the system was to illustrate the capabilities and potentialities of computerized reference search systems. The system's purpose was not to serve as a perfected retrieval system ready to use at the patron level, although it certainly could, but instead to serve as a prototype retrieval system that could be expanded and further refined for practical use at the reference desk.

Delimitation of the Problem

The system was limited in scope. The data base was only composed of biographical reference works regarding persons who live within, or who have lived within, the United States from Colonial times to the present. Furthermore, it only included works presently held by the Harold B. Lee Library General Reference Department and shelved on the main (third) floor of that library. All other biographical works on

other floors were not included. Although this retrieval system was "on line" and fully automatic, the assistance of a qualified person was still necessary to formulate and enter the patron's search profile into the system. However, it would be possible to augment the system with computer-assisted instruction, thus enabling the system to function with a minimum of supervisory assistance.

The study did not discuss computer programming techniques or file organization to great length. Moreover, the study did not deal with philosophical or theoretical concepts of reference service, nor with philosophical concepts of information theory, information science, or computer-assisted instruction systems or techniques.

CHAPTER II

LITERATURE SEARCH

Attempts to automate library processes, which began seriously after World War II, have focused mainly on technical services including cataloging, acquisition, circulation control and serials. Little has been done in the area of reference work to utilize the capabilities of the computer. Although the reference librarian has seldom directly benefited from automation, he has benefited indirectly. For example, in some libraries the holdings are on magnetic tape or in "machine readable form" which makes it possible to produce both massive and selected bibliographies on given subjects. "Key word in context" as well as "citation" indexes are computer produced and established to provide bibliographical control more effectively for the reference librarian.

Except for the two studies discussed in this paper, there are very few instances of functional automated systems which give direct and immediate help to reference queries. There have been some attempts to streamline bibliographic control of reference works using card systems. These include the well known "punched card," "peek-a-boo," "edge-notched," and "uniterm" systems.[1] Each of these attempts to act as a means by which sources are categorized and subdivided on a card and then visually

---

[1]Claire K. Schultz, "Automation of Reference Work," Library Trends 12 (January 1964): 413-14.

examined by the librarian to determine appropriate sources for a particular need. Although they offer some help, they are not very practical or efficient, nor do they use the speed and memory qualities of a computer.

Two research projects, described in the literature, report experimentation with automated reference search systems. These projects were conducted in the last five years; one at the University of Chicago and the other at the University of California at Berkeley. Both of these projects attempted actually to automate a reference system. Automation, in the context of these two experiments, means the indexing of reference works with descriptors, the feeding of this information onto a rapid access storage device, and the organizing of a data bank for retrieval purposes. To date, these have been the only two major projects dealing with computerized reference services. Much of the remainder of the literary search portion of this paper was devoted to an examination of these two systems, giving a good idea of the state of the art of automated reference systems.

Examination of Weil's System

The first system examined was the one established by Miss Cherie B. Weil at the Graduate School of Library Science at the University of Chicago in 1967. Miss Weil attempted to establish an automated biographical reference search system that would advise the patron which types of biographies would be most appropriate to check in terms of a given biographical reference question.

Weil assumed that biographical reference works (which are confined to information about people) have something in common. She felt that

most biographical works generally give the same type of information for each person.[1] This assumption further led her to assume that it was possible to index biographies within a computer in such a way as to establish an automated retrieval system. Mechanical retrieval could be achieved by matching the index terms assigned to the reference works with the patron's request through an appropriate retrieval rule. Weil felt that an advantage of such a system would be to free the librarian so that he could engage in less mechanical and more scholarly tasks.

Weil's system dealt with all types of biographical reference works from all types of occupations and nationalities. These biographies were indexed in terms of a unique classification scheme. She used nine different index descriptors to describe each biographical work. They are:

(1) Living (Are the subjects considered in the book living or dead?)
(2) Nationality (Does the work include only certain nationalities?)
(3) Occupation (Is the work limited to only certain occupations?)
(4) Minority Groups (Is it limited to specific minority groups?)
(5) Date (What time period does the work cover?)
(6) Indexes (What separate indexes, if any, index the work?)
(7) Sex (Is the book limited to any one sex?)
(8) Specifics 1 (Other information that describes the work.)
(9) Specifics 2 (Other information that describes the work.)[2]

The following is an example of a biographical reference work that is completely coded:

DICTPHILBIO/ INDEX FIELD X, LIVING N X, OCCUP Z, SEX A, NAT PHILP ASIAN X, SPEC 1 D C DS FL BP L CL CM DG E I Z, DATA 50's X, SPEC 2 P PL R MS PD Z, MIN Z+.[3]

Interpreted, the above entry means: Dictionary of Phillippine Biography;

---

[1] Cherie B. Weil, "Automatic Retrieval of Biographical Reference Books," Journal of Library Automation 1 (December 1968): 240.

[2] Ibid., p. 241.

[3] Ibid., p. 242.

a book limited to dead Filipinos and giving for each entry, the dates, career, descendants, field, birthplace, long articles, class in college, degrees, education, pictures, parents, publications, references, marital status, and physical description. Each reference work was indexed in terms of the nine divisions, since a fixed format was used in the file structure. This meant that a certain amount of space was allotted each index category for each work indexed.

Weil indexed, in the above manner, many different biographical reference works. The data were stored on tape, disk or cards, which would then be accessible by computer, thus establishing a sizeable computer data bank of biographical reference works. This data bank was then used to benefit both the librarian and the patron directly. For example, if a patron wanted a biography of a Dutch economist, that question could be fed into the machine and those works whose index records matched the request would be retrieved.

The reader may ask the question: "How was the machine able to interpret the patron's request for a biography of a Dutch economist?" This request was coded in the same manner as the biographical reference works were coded, so that the request for a Dutch economist would look like the following:

> Question: SEX Z (or M), LIVING Y, NAT DUTCH, OCCUP ECON, MIN Z, DATE 60's, SPEC 2 PL+.[1]

It was especially important that the patron's search profile be coded using exactly the same descriptors with which the works were indexed. Furthermore, the search profile must be constructed in the same order in which the works were coded. This means, for example, that the

---

[1] Ibid., p. 243.

category "SEX" should always be first in the search profile, the category "LIVING or DEAD" should always be second, the category "NATIONALITY" whould follow third, and so forth. In other words, for a match to occur the search profile must be in exactly the same order as the works were indexed.

As can be seen from the above search profile, the machine is programmed to compare the coded abbreviations of the patron's request with the coded abbreviations of the biographical reference works within the data bank. The computer simply prints out the results of the match. This is basically how any automated reference system works.

Examination of Meredith's System

Following Weil's research in 1967 was another research project started in 1970 by Mr. Joseph C. Meredith at the Institute of Library Research at the University of California at Berkeley. Meredith's research was funded by the Office of Education. Mr. Meredith had a larger staff, more funds, and was more sophisticated in establishing a larger data bank, but implemented some of the same elements as Weil. The two projects were similar in some respects, but in others, differed greatly.

Meredith's system initially began as an attempt to construct a computer-assisted instruction system, not a reference search system. However, in the process of the former, the latter developed. His REFSEARCH system, like Weil's, retrieved reference sources, not direct answers to the patron's question. As Meredith put it, "It retrieves sure sources of information which are likely sources of specific information."[1]

---

[1]Joseph C. Meredith, Reference Search System (REFSEARCH) Users' Manual, (Berkeley: Institute of Library Research, University of California, Berkeley, 1971), p. 1.

His entire collection was divided into approximately sixteen different major divisions. In turn, each division was then subdivided by five different sets of subdivisions. Ultimately, the limitations of the system's usefulness depended on the exactness of the classification scheme. Suffice it to say that Meredith established a means by which he could identify and describe his reference collection.

As mentioned above, Meredith's REFSEARCH system was much more comprehensive and sophisticated than Weil's. Meredith's staff spent about ten and one-half man-months of labor before the entire coding or indexing of his 790 different reference works was completely finished.

Not only was Meredith's system more extensive, it differed from Weil's in that the former assigned numerical codes to each of his descriptors. For example, in Meredith's system, the descriptor "LOCATES" could be represented by the numerical code 345, or the descriptor "PROPER-NAMED" could be represented by the numerical code 334 and so on. The reason for this is that although the computer can handle and match series of letters, it is more efficient, speedy, and economical when it deals with numbers. Furthermore, when alphabetical symbols are used, then an "ASSEMBLER" must be written to translate the characters to numbers. To communicate in numbers is less convenient to the user, but more efficient for the computer.

In Meredith's system, patron queries are submitted as a series of numbers separated by commas. An example of this will help to clarify how his system worked. Suppose a patron came to the reference desk with the following question: "I've heard that Kurt Godel, the mathematician, lives somewhere in the United States. Where can I find his home address?" The librarian then studies the question, determines what kinds of

information are being dealt with and what kind of service is needed, and then proceeds to express the question in terms of the aforesaid descriptors. Hence, in terms of Meredith's descriptors, the patron's question might look like the following:

> Required: A work which LOCATES (345) REAL (339) PROPER-NAMED (334) INDIVIDUAL (335) LIVING (341) PERSONS (333). Must cover MATHEMATICIANS (528) in the UNITED STATES (558).[1]

The librarian would then type the following numbers into the computer: 345, 339, 334, 335, 341, 528, 558. The computer compares the set of numbers with the numbers within its data bank, makes a match, and then prints out the works that matched. Except that Meredith's system used coded numbers to describe reference works, while Weil's used words to describe reference works, the two systems employ the same basic elements and procedures to implement the system.

Summary and Conclusions of Literature Search

The foregoing pages have attempted to analyze and describe the literature which describe what was involved in implementing a reference search system. The literature search has especially centered on two research projects that explicitly dealt with the implementation elements involved in a reference search system. The first project was Cherie Weil's and the second was Joseph Meredith's. Although the two projects were independent of each other, some of the same elements and procedures were followed to implement both systems. These two projects were the only ones in the literature that have seriously attempted to automate reference services.

---

[1] Joseph C. Meredith, "Machine-Assisted Approach to General Reference Materials," Journal of the American Society for Information Science 22 (May-June 1971): 182.

Although Weil and Meredith actually constructed a computerized system that worked, neither of them used it at the reference desk level, so the library patron did not directly benefit from it. Both systems were used as an experimental prototype from which future systems could conceivably be built and used at the public service desk.

Both systems used a unique classification scheme to index the reference works; however, neither Weil nor Meredith claimed these schemes were fool proof. On the contrary, they both admitted the system was no better than their application of the index terms to their reference collections. Indeed, Meredith admitted that his descriptors were not consistently applied throughout his collection. Meredith also said that any classification scheme is to a certain extent artificial in its attempts to identify and distinguish one reference work from another.

Another criticism of Weil and Meredith was that initially neither system was capable of ranking the reference sources, from which a patron could find an answer to his question, in order of importance. For example, suppose a patron asked the question: "Who is Jesse Shera?" The system might list two reference sources from which an answer could be found, namely; Who's Who in America and International Who's Who. As far as the patron is concerned, both sources equally possess the same amount and kind of information on Shera, when in reality the former, Who's Who in America, would probably contain more information on him.

One of the remarkable qualities of Meredith's and Weil's systems was the fact that they both harnessed the speed, memory, and efficiency capabilities of the computer. This machine can very quickly compare any combination of terms, arrive at a "match," and then print out the

works that deal with the combinations of topics desired. This is an advantage over card catalog systems where only one topic at a time can be dealt with efficiently. In terms of speed and accuracy, although never officially and unequivocally proven, the computer certainly has the potential for superiority over the human librarian, since it never forgets any of the hundreds of details associated with each of the hundreds of reference books.

Both Weil and Meredith followed a definite procedural pattern in establishing their systems. Certain sequential steps had to be followed in order to make their systems work. The most valuable aspect of the literature search portion of this paper was in studying the procedural steps taken both by Weil and Meredith. These steps were followed by the author as he attempted to establish an automated reference search system. These steps were outlined in the methodology statement within the paper.

The literature search chapter yielded some excellent, scholarly, in-depth information on certain research projects conducted on the subject of automated retrieval systems. These projects, although imperfect, provided excellent information on the state of the art and on the implementation of a computerized reference system. These projects pointed the way into automation of public services, which is far behind automation of technical services. The projects are admittedly experimental, to a certain extent unrefined, and untried in terms of actual patron use, but they have given the author a basic formulation and understanding of computerized retrieval systems.

CHAPTER III

BUILDING THE RETRIEVAL SYSTEM

Methodology

An automated reference system was created by following basically the same procedural steps as Weil and Meredith. These steps follow:

(1) The system's "design" limitations were established. These limitations prevented the system from "growing" too large because of the nature of file organization, memory size, and the like. Basically the exact types of works and how many of them are to be dealt with in the system were decided during the preliminary stages. This was important since these initial boundaries limited the kinds of requests which were acceptable to the system and determined the areas with which the computer dealt. The system contained only biographical reference works that had information about persons who are living or have lived in the United States, including Alaska and Hawaii, from Colonial times to the present in all areas of activity. The system was limited to only those biographical works housed or shelved on the third (main) floor General Reference area of the Harold B. Lee Library. All other biographical works on other floors were not included.

(2) After the basic limitations were set, the system's scope was further limited by devising descriptive surrogates for each biographical reference source. The definite level of specificity and depth to which the surrogates or descriptors would be applied was also established. In other words, a unique descriptive classification scheme for indexing

the works was prepared. This step was most important, since this classification scheme determined the extent to which the system was able to retrieve sources effectively. The author indexed the biographical reference works in terms of (1) a miscellaneous category set, which included sex, regional location within the United States, religion, race, nationality within the United States, and age, (2) occupation categories and (3) time period categories. All works were indexed in terms of these three areas.

(3) An authority list of the descriptors or index terms was compiled for use in the classification scheme. This list provided a means of standardization when indexing the works and also provided the base into which all requests were converted.

(4) The works were indexed in terms of the classification scheme by using the descriptors or index terms to describe and identify the works. Care was taken to apply accurately and consistently these terms throughout the collection of works.

(5) The information was stored on computer disk. This data base was developed by directly typing in the information through the computer terminal. Information included all the bibliographic data in the works themselves (such as title, call number and explanatory notes) as well as their respective descriptors. The information may be stored not only on disk, but on any machine-readable medium such as cards, magnetic tape, paper tape, drums, etc. Moreover, it is not mandatory that the data be encoded by using only the terminal. For example, the data could be keypunched onto cards and the cards then read into the computer. The key point is that bibliographic information must be converted to machine-readable form.

(6) The software was created, which instructed the computer how to match the patron's question with its data bank of biographical works. The author chose to write the program in COBOL language under the direction of Mr. Bruce Smith, Mr. Don Novak and Mr. Larry Gibson, all of whom are recent graduates of the computer science department at Brigham Young University. Although the COBOL language was used and is very efficient in dealing with strings of characters, it is not the only possible language that could have been used for this project.

(7) The patron's questions were expressed in terms of the unique descriptor terminology. This enabled the computer, by using standardized index terms, to match the patron's question with its extensive data bank of biographical information. The librarian must convert the patron's natural language question to a logical descriptor language question by using the list of index terms as developed in step three.

By carefully following the above methodology, the author established an experimental computerized reference system that accurately retrieved sources of information. This system can be used to explore potential improvements to user service at the reference desk.

The remaining portion of this project applied each of the steps listed in the methodology statement to the building of the retrieval system. Since this project focused around building an automated retrieval system, and since the methodology specifically dealt with that, the remainder of the project followed the methodology statement as a sequential outline.

The author chose to deal with only United States biographical reference works that were shelved in the general reference area of the third floor of the library. Over 170 such titles, consisting of many

more volumes, were shelved in the general reference area. The author chose to deal with biographical reference works for several reasons: first, there was no exhaustive index that provided bibliographic access to the thousands of names listed in the hundreds of works; secondly, there were only a few reference tools, such as Winchell's Guide to Reference Works and Slocums' Biographical Dictionaries, which categorized biographies by occupation. These latter works provided a service similar to the computerized retrieval system, but the retrieval system had other advantages; thirdly, biographical works contain basically the same type of information, are arranged in similar formats, can be indexed in terms of a general classification scheme, and hence, lend themselves to this sort of project.

Perhaps the most important step in this entire project was to devise an extensive, unique classification scheme which was appropriate to biographical reference works and which was both general as well as somewhat specific in nature. "General" meant, very broad, widespread categories such as "scientist" or "businessman." "Somewhat specific" did not mean explicitly precise, but refered to an intermediate level of specificity, such as the terms "chemist" or "accountant." Explicit precision, in the classification scheme, was unnecessary, since most biographical reference works only reached the intermediate level of specificity. Therefore, unless stated otherwise, the word "specific" did not refer to explicitly precise but only to an intermediate level of specificity.

An explanation of the importance of an accurate classification scheme was necessary. A document was indexed, being given the descriptors, "evaluation," "library schools," and "curriculum."

The implication was the document dealt with the evaluation of library school curriculum, but later the document was actually found to deal with the evaluation of elementary school library programs. In this case, the assigned descriptors did not accurately describe and identify the subject of the document. A user would find it both disappointing and annoying to retrieve such a document only to find it unrelated to his desired topic. Thus, any classification scheme must contain descriptors which are precise in meaning and succinctly descriptive. These descriptors must also be applied with equal consistency over the entire data base, as discussed later in this paper.

An analysis of the essential areas of information to be described in biographical reference works was necessary before a classification scheme could be devised accurately. The analysis found essentially three special areas which should be emphasized in biographical works to describe them accurately. The three areas were: (1) Miscellaneous categories, which included regional location within the United States, sex, religion, race, nationality within the United States, and age; (2) occupation categories, and (3) time period categories. These three areas adequately and yet simply described most biographical reference works which contain information about people within the United States both from Colonial times to the present and in all fields of activity.

This scheme is outlined on pages eighteen, nineteen and twenty of this paper. Pages nineteen and twenty illustrate the various occupation areas and their respective symbols. There are eighteen areas which are subdivided into other occupation areas. The large majority of biographical reference works can be properly and accurately indexed,

Descriptor Classification Scheme

Field One- Miscellaneous Categories

Chart no. 1

A SEX

- B MALE
- C FEMALE

D UNITED STATES

- E EAST (includes Middle Atlantic and Northeastern States)
- F SOUTH (includes Southern and South Central States)
- G MIDWEST (includes Central and Midwestern States)
- H WEST (includes California and the Western States)
- I NORTHWEST (includes the States of Oregon and Washington)
- J ALASKA
- K HAWAII

L RELIGION

- M MORMON
- N CATHOLIC
- O PROTESTANT
- P JUDAISM
- Q OTHERS

R RACE

- S NEGRO
- T INDIAN
- U ORIENTAL
- V OTHERS

W NATIONALITY

- X EUROPEAN
- Y SCANDINAVIAN
- Z ASIAN
- Ø AUSTRALIAN
- 1 ISLANDS
- 2 NORTH AMERICAN (includes Mexico, South America and Canada)

AGE

- 3 Ø to 18 years old
- 4 19 to 35 years old
- 5 36 to 110 years old

NOTE: The Classification Scheme will accomodate growth of new subdivisions, simply by adding new alpha-numeric symbols. For example, the religious category, "Buddhism" could be represented by the symbol "P_1", or the occupation category "Management", by the symbol, "J_1". In this way the Classification Scheme can be further expanded and subdivided if necessary, simply by inserting these symbols in the appropriate position within the scheme.

Descriptor Classification Scheme

Field Two - Occupation Categories

Chart no. 1

\+ ALL OCUPATIONS

A ARTISTS
- B PAINTING
- B ENGRAVING
- B SCULPTURE

- C MUSIC
- C THEATRE
  - C MOTION PICTURES
  - C RADIO and TELEVISION
  - C ENTERTAINMENT

- D ARCHITECTS

E ATHLETICS
- E SPORTS AND GAMES

F AUTHORS
- G POETS
- G NOVELISTS
- G PLAYWRIGHT
- G JOURNALISTS

- H CHILDREN WRITERS

I BUSINESSMEN
- J PHILANTHROPISTS
- J BANKING AND FINANCE
- J INDUSTRY

- K ACCOUNTING
- K STATISTICS
- K ECONOMICS
- K ADVERTISING
- K PUBLIC RELATIONS

L EDUCATORS (includes all levels and all disciplines)
- M LIBRARIANS
- N PROFESSORS (includes all disciplines)
- O ELEMENTARY AND SECONDARY EDUCATORS

P ENGINEERING AND TECHNOLOGY
- Q MATHEMATICS
- Q COMPUTER SCIENCE AND DATA PROCESSING

- R AVIATION

- S ATOMIC SCIENCE

BEST COPY AVAILABLE

Descriptor Classification Scheme

Field Two - Occupation Categories (con't)

T EXPLORERS Chart no. 1

U HUMANITARIANS (social and economic reformers)

V HUMANITIES
- ENGLISH - See EDUCATORS
- HISTORY - See EDUCATORS
- LITERATURE - See AUTHORS
- (PHILOSOPHY - See Also EDUCATORS
- W LINGUISTICS - See Also EDUCATORS

X INVENTORS

Y LAW

Z MILITARY SCIENCE
- Ø ARMY
- 1 NAVY
- 2 AIR FORCE

3 MEDICINE
- 4 PHYSICIANS AND SURGEONS
- 4 NURSES
- 5 DENTISTS

6 PHYSICAL AND BIOLOGICAL SCIENCES
- 7 ECOLOGY
- 7 BOTANY
- 7 AGRICULTURE
- 7 AGRONOMY
- 8 ZOOLOGY
- 8 BIOLOGY
- 8 ANIMAL SCIENCE
- 9 CHEMISTRY
- ; PHYSICS
- ; GEOLOGY
- . FOOD SCIENCE

/ SOCIAL AND BEHAVIORAL SCIENCE
- : SOCIOLOGY
- ' POLITICAL SCIENCE
  - ' STATESMEN
  -) PRESIDENTS OF THE UNITED STATES
- ? ANTHROPOLOGISTS
- " ARCHEOLOGISTS
- $ GENEALOGISTS
- % GEOGRAPHY
- - PSYCHOLOGY

\- STUDENTS (includes secondary and higher education)

\# THEOLOGIANS

as far as occupation is concerned, by this set of occupations.

Page eighteen illustrates a variety of index areas, including sex, regional location within the United States, religion, race, nationality within the United States and age. As the author indexed the 170 various works, he found it necessary to index them not only in terms of occupation and time period, but in the areas of regional location, sex, religion, race, nationality and age as well. For example, how does one accurately describe such biographical works as, Negro Ministers, Who's Who of American Women, Who's Who of Polish America, Who's Who of Scandinavian America, Who's Who in the Catholic Church or Who's Who Among High School Students? If the index areas of occupation are used along with time period, the above mentioned works could not be accurately described, and hence, they could not be retrieved properly. In other words, these works need more than just the two index areas of occupation and time period to describe them accurately. The index areas, as illustrated on page eighteen supply this need.

Because of the limitations of the computer program, the areas as outlined on page eighteen could not be intermixed within themselves, although they could be intermixed with the occupation areas as outlined on pages nineteen and twenty. For example, the areas such as religion, race and age cannot be mixed together with scientists; only religion and scientists, or race and scientists, or age and scientists can be mixed together to describe a reference work. This means that a work entitled, Catholic Negro Scientists can only be indexed as Negro-scientists or Catholic-scientists but not Catholic-Negro scientists. The work can be indexed in such a way, however, that if a patron searched for either Catholic-scientists or Negro-scientists he would

BEST COPY AVAILABLE

retrieve the work Catholic-Negro Scientists. In other words, the work is indexed under both index terms, but not both at the same time. This limitation is not serious in terms of patron satisfaction, but it does mean the user will, under some situations, have to pass twice over the data base to be assured of getting all the works possible.

A few other explanations need to be made concerning the descriptors on page eighteen. The heading SEX is used to describe such works as Who's Who of American Women, where the work deals with one sex exclusively. The heading UNITED STATES was divided into regional areas. The EAST, for example, denoted all states lying within the area designated as the Middle Atlantic and Northeastern States. A regional index of terms was needed to describe such works as Who's Who in the East and Who's Who in the West. Other works such as Who's Who in California and Who's Who in Oregon needed to be indexed with a regional location descriptor. The latter two were indexed with the descriptor WEST, since that is as specific as the index terms will allow.

The heading RELIGION was subdivided into very general denominations, since there are very few biographical reference works which are more specific in format. This is also true for regional location biographical works. There are not many states, for example, which publish their own biographical works. In other words, there is no need to particularize a classification scheme with numerous specific index terms which are very rarely used. It is better to have what could be called "intermediate" index terms which lie between the very general ones and the very specific ones.

Another heading on page eighteen is RACE. This heading was divided into general, but specific-enough, sub-index terms. The heading

NATIONALITY is similarly divided into areas of the world. A reference work entitled, Who's Who in Polish America, received the index term EUROPEAN, while the work entitled, Who's Who Among Japanese Americans, was indexed by the term ASIAN and so on. The last heading on page twenty is AGE. The "0 to 18" category is used to index a work such as Who's Who Among High School Students.

There are eighteen major occupational categories, each one subdivided into various areas. The ALL OCCUPATIONS category included all the other seventeen categories. The BUSINESSMEN category was divided into eight subdivisions. The author has not tried to be all inclusive with regard to these subdivisions, but has only attempted to divide the main headings into "intermediate" subdivisions so that the reader may know generally what occupational categories come under which headings. As mentioned earlier, there is no need to be very specific since most biographical works tend to be general or "intermediate" in scope.

The time period category was not outlined in the "Descriptor Classification Scheme" because no set time periods were used. If work dealt with people living from 1901 to 1965, those years were used to describe the work. If a work was published in 1936 but was retrospective and covered persons living from Colonial times up to 1936, then the dates 1600 to 1936 would be used. Most works usually state what time period they cover and in these cases it is very simple to affix a set of beginning and ending dates. However, there are a considerable number of works that do not plainly specify the time period covered. A work entitled, Who's Who in the Rockies, for example, was published only once, in 1923. It was not retrospective; that is, it only dealt

BEST COPY AVAILABLE

with persons living at the time it was published. Therefore, the beginning and ending dates 1923 to 1923 were assigned to it.

To assign only one year to a biographical reference work seemed too limiting. For example, if a patron wanted a work on people in the Rockies in 1922, then Who's Who in the Rockies would not be retrieved since the date 1922 does not match the date 1923. At first examination this seems a distinct disadvantage to the user. If he does not want the precise date underwhich the work is indexed, then there will not be a match and the work will not be retrieved. Therefore, the computer program was written so that the user may or may not include a date in his search. This meant that a patron may use only the first two groups of index areas (miscellaneous and occupation categories) in his search expression without including the time period category.

It would certainly be possible, although it is not presently built into the system, to alter the computer program so that the user could include in his search profile the expression $\leq$1923, which means less than or equal to 1923. This expression would ensure that the user could retrieve all the relevant works within the data base that were equal to or less than the year 1923. However, since this option is not possible within the present system, the user will have to be satisfied with the "no date" search routine as explained above, which is almost as satisfactory as the "equal to or less than the date" routine. More will be said later on about retrieval rules and patron search profiles.

On pages eighteen, nineteen and twenty, there were certain symbols assigned to each descriptor. The symbol A, for example, on page eighteen means SEX; the symbol S means NEGRO, and so forth. Both

the main headings as well as the sub-descriptors have symbols. The biographical reference works, within the data base, were indexed in terms of these symbols. The symbols on page eighteen are used again on pages nineteen and twenty. This does not matter, since, to the computer, the symbol S on page eighteen (which means NEGRO) is entirely differentiated from the symbol S on page nineteen (which means ATOMIC SCIENTIST).

To reiterate, assigned to every biographical reference work within the data base are three fixed fields or categories; one field for miscellaneous descriptors, as found on page eighteen, another field for occupation descriptors, as found on pages nineteen and twenty, and a third field for time period or dates. To the computer, each of these fields is separate and distinct, and it does not matter if the same symbols are used twice, as long as they are in different fields. It would matter; however, if the symbol S, for example, were used twice in the occupation categories. More was said regarding these fields in the discussion of programming techniques and file organization.

The descriptors on pages eighteen, nineteen, and twenty not only serve to illustrate the author's classification scheme but also serve as an authority list or subject headings list. This is vital, since the descriptors, as well as their symbols, must be applied or assigned with equal consistency and accuracy over the entire biographical reference collection. This is basic to all information retrieval systems. Both the ERIC Thesaurus and the Library of Congress Subject Heading List require the same constraints.

With the classification scheme established and refined somewhat, the next step was to index all the biographical reference works in the General Reference Department of the Lee Library. All of the 170 works

were individually examined to determine their contents. They were then indexed in terms of the three categories previously explained.

Appendix A is a partial list of some of the biographical reference works in the data base. Before each of the titles appears a series of symbols such as HDF or '/#. These symbols, as explained, were used to signify the descriptors on pages eighteen, nineteen and twenty. In Appendix A, all of the titles were indented evenly. There was a good reason for this. There were, as mentioned, three categories in which all works were indexed. Each of the three categories had a fixed field format; that is, each category was allotted so many spaces within the fixed fields. The first fixed field (miscellaneous) was allotted five spaces from the left hand margin or the beginning of the record. The second fixed field (occupation) was allotted seven spaces and followed directly after the first fixed field. The third fixed field (time period or dates) was allotted eight spaces and immediately followed the second fixed field. The following example will help to clarify this arrangement:

D /' 1774l971 BIOGRAPHICAL DIRECTORY OF THE AMERICAN CONGRESS 1774-1971 Ref 923.2 UN332

The D , followed by four spaces, is the first fixed field and comprises the miscellaneous category. Only one of the five spaces is filled with symbols, namely the D. The /' followed by five spaces is the second fixed field and comprises the occupation category. Only two of the seven spaces are filled with symbols, namely the / and the '. The 1774l971 is the third fixed field and comprises the time period category. All eight of the spaces allotted to this category are filled and always will be filled, even if the same date appears twice, such as 19231923. Since a fixed field format was used, it is

BEST COPY AVAILABLE

unnecessary to have spaces between the dates, or for that matter between the three fixed fields. The reference work, as illustrated above, is indexed to mean: a biographical work containing biographies of statesmen living within the United States between 1774 and 1971.

After the third fixed field (date category) there appears such information as the title, the dates the work covers (repeated again), the reference call number, and if useful, explanatory notations. There are 95 spaces allotted from the date category to the end of the record for such information. The notations (por) mean that the work contains portraits of its biographees. All letters were capitalized in the data base, including cutter letters, since most computer processing deals exclusively with upper case letters.

The following example illustrates another important factor about the way in which the works were indexed:

LP +LZ/#119551965 WHO'S WHO IN WORLD JEWRY 1955-1965 R 922.96
W62S

In the first fixed field, the letters L and P appear, which signify the descriptors RELIGION AND JUDAISM. In the second fixed field, the symbols +LZ/'#1 appear, which mean respectively, ALL OCCUPATIONS, EDUCATORS, MILITARY SCIENCE, SOCIAL SCIENCE, STATESMEN, THEOLOGIAN, and ARMY. The work was indexed by both the general and "specific" (meaning intermediately specific) descriptors, so that a patron would be assured of retrieving the work whether his approach was general, such as RELIGION, or a little more specific such as JUDAISM. Whenever a work was indexed by a sub-category such as JUDAISM, ARMY, and STATESMEN it was also indexed by its more general heading RELIGION, MILITARY SCIENCE, and SOCIAL SCIENTIST. Thus, the work was assured

of being retrieved whether the patron searched under general or more specific terms.

The double indexing also allows special bibliographies to be printed out both for specific as well as general areas. The above work, Who's Who in World Jewry, could be retrieved by inputing the following combinations of symbols:

L+ or LL or LZ or L/ or L' or L# or Ll or P+ or PL or PZ or P/ or P' or P# or Pl.

The fact that each work was indexed under not just one descriptor from each field, but rather, under many descriptors from each fixed field, makes the system highly practical.

The author himself indexed the entire collection of 170 works. Each work was indexed as specifically as possible, but whenever a work was indexed by specific categories it was also indexed by its more general heading.

Now that the works were indexed and checked for accuracy and consistency, the information was stored onto disk through the computer terminal. Each of the descriptor symbols as well as the bibliographic information was individually typed into the data file designated as "Books.fil." Corrections were made using the TECO (text editor correction) routine. It was important to maintain the exact format, as described earlier, as the works were inputed. It took the author about seventeen hours to input the data, make corrections, and establish the data base. This data base was partially reproduced as Appendix A of this paper.

The computer program was the last remaining essential to the project. It was written under the direction of Mr. Bruce Smith, Mr. Don

Novak, and Mr. Larry Gibson, all of whom are computer science graduates or students. The program's meaning, limitations and capabilities were fully explained in Appendix B. In essence, the program instructs the computer how to match the patron's profile with the index terms assigned to the biographical reference works. The program was written using the COBOL language, which was especially able to deal with long strings of characters, as opposed to numbers on the other hand.

This program was quite complex since the nature of the task was complex. The machine must be told in precise detail every step of the search routine. It will not do anything it has not been instructed to do. However, the computer has tremendous memory, speed, and accuracy capabilities. It was able to compare the search-key with each of the 170 works and make a match within milliseconds. The user, sitting at the terminal, usually spends only seconds before the computer responds and displays works retrieved.

Another item of importance was the search-key; which is composed of symbols that are typed into the computer just before the execution of the program. It represents the patron's question in terms of the unique descriptor classification scheme. It was essential for the user to convert his information need into a set of terms which may be used as a search-key. This was done by using the descriptors listed on pages eighteen, nineteen and twenty. The following have been converted to symbolic language search-keys:

I would like a biography of Hank Aaron, could you help me find one?
This was converted to, SE1974 (Negro Athletics in 1974)
Or DE1974 (U. S. Athletics in 1974)
Or DE (U. S. Athletics)

I would like a biography of Eugene O'Neill, the famous playwright.
This was converted to, DG1955 (U.S. Playwright in 1955)

BEST COPY AVAILABLE

It was also converted to DF1955 (U.S. Author in 1955)
Or DF (U.S. Author)

I would like a biography of James Chipman, a field Zoologist, who also teaches at the University of Iowa.
This was converted to, D8197h (U.S. Zoologists in 197h)
Or D6197h (U.S. Physical or Biological Scientist in 197h)
Or DN197h (U. S. Professor in 197h)
Or G8197h (Midwestern Zoologists in 197h)
Or G6197h (Midwestern Physical or Biological Scientist in 197h)

In each of the above examples, the user went from the specific to the more general search-key to ensure retrieval of all works that are relevant to the question. The user may or may not search with a date, which also facilitates retrieval of all possible works. Since a match must occur on only two categories, namely miscellaneous and occupation, and not the third category of time period, very few works are not retrieved that are relevant to the patron's question. However the user must determine in what categories the patron's biographee falls. Hank Aaron, for example, is a famous Negro baseball player; however, not all biographees are as well known. Sometimes, the librarian must question the patron about which miscellaneous, occupational and time period categories fit the biographee.

It was important to convert the question into the proper symbols, since the works themselves were indexed in terms of those symbols. The whole retrieval system was based on the matching of the search-key symbols with the symbols that indexed the biographical reference works within the data base. Moreover, for a "match" to have occured, and for works to be retrieved, all elements must be matched. This meant that if the search-key appeared as, DE197h (U.S. Athletics in 197h), all three elements must be matched before any work can be retrieved. In other words, the D and E and 197h must all be matched. This is similar to

the "anding" of words together in other retrieval systems.

Appendix C illustrated several searches actually made within the system and demonstrated its versatility. These searches, in Appendix C illustrate different combinations of descriptors which yield different results. When the date is left out of the search-key, the number of reference works retrieved increases. Several subject bibliographies produced by the system appeared near the end of Appendix C. This in itself provides a valuable service to students and teachers alike. For more information on the search-key, the details of the computer program, as well as the ability of the user to search both with or without a date, Appendix B was written.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The problem, briefly restated was to construct an automated retrieval system for biographical reference works. This retrieval system, although limited in scope, was to serve as a prototype of future retrieval systems and to illustrate the capabilities and potentialities of computerized reference retrieval systems.

Such a prototype system was indeed successfully designed and implemented. The system was based on the same methodology used by Weil and Meredith in their systems, thereby giving credence to the statement that indeed there are some universal or at least general principles and procedures to follow in building an automated system. However, it was apparent, during the course of the project, that much research and further study needed to be undertaken in the area of reference automation before such a system can be ideally used at the patron service desk. The main advantages of computerized systems over the conventional system are the following:

(1) The memory, speed, and accuracy capabilities of the computer can be utilized to their full extent.

(2) With the availability of leasing computer time from service bureaus and with access to time sharing computer systems, the usage of heretofore expensive equipment is now becoming cost effective.

(3) It is very easy, with the capabilities of the computer, to combine various descriptors which are both general and specific, whereas

BEST COPY AVAILABLE

the conventional reference works, such as Winchell and Slocum, can only index biographical works under one descriptor at a time. The computer is easily able to combine any number of terms and search for a match.

(4) Special bibliographies on various subjects are easily produced with great speed and accuracy using the computer. This in itself is valuable to teachers as well as students.

(5) Since reference works are already on disk in machine-readable form, for the purpose of reference retrieval, it is also possible to use the data base to help select new materials. The librarian, with the aid of the computer, can readily spot deficiencies within the present collection.

(6) The ultimate and most important usefulness of reference automated retrieval systems is to the patron. Hopefully, in the not-too-distant future, entire reference collections can be in machine-readable form. Only time and more research stand in the way of transforming the world of reference work and libraries into more efficient, speedy, and even economical data and information retrieval systems.

With the foregoing elaboration of the advantages of retrieval systems, it is appropriate to enumerate also the disadvantages, limitations, and problems associated with developing such systems:

(1) Any automated reference retrieval system, no matter how complex and elaborate the hardware and even the software, is only as good as the classification scheme that is used to index the works. The problems associated with "surrogates" and their respective ambiguity applies here. The whole notion of what is relevant and what is necessary to represent through surrogates appears as an ever puzzling problem. What one person, such as the indexer of reference works, may think is relevant, may be

completely different from what the patron believes is relevant. In other words, the computer and the software can only judge the relevance of a reference work based upon the index classification scheme, which may itself be relevant to one person but inaccurate and irrelevant to another.

(2) Associated with the first disadvantage, is the problem of applying the index classification scheme to the collection of reference works. Even indexing the collection by one person, as was the case with this project, does not ensure that the descriptors have been applied consistently. On the contrary, even when one person indexes the works, much variation occurs.

(3) The talents of an "intermediary" (be it a professional reference librarian, or a trained library assistant) are necessary to question the patron, determine his needs, formulate a search-key profile, and input it into the computer.

(4) Although the system was only intended to be a prototype of future retrieval systems, much can be learned through user studies. At present, no automated retrieval system, including the author's, has been very extensively used at the reference desk level. Certainly, this is a major milestone to be accomplished in the future.

(5) At present, the author's system was structured so that the matching routine was accomplished using a sequential search. This means that the computer has to compare the search-key with all of the works in the data base one at a time, starting from the beginning and going to the end of the file. More experimentation is needed to determine the optimum file organization for increasingly larger and larger data bases, since sequential searching routines can be both inefficient and expensive

when dealing with large data bases.

(6) As mentioned earlier in this paper, Weil's, Meredith's and the author's systems, were not able to rank the reference sources in order of importance. Ideally, those works most likely to give the patron a correct answer would appear first, while less likely sources would appear later.

(7) For automated systems to become completely realistic, they must be able to deal with more sophisticated, philosophical questions and not only simple, data retrieval, type questions.

It is recommended that additional research projects be devoted to experimentation with reference retrieval systems, and especially to the building of classification schemes, which ultimately limit the capabilities and are the basis of any retrieval system. If constant research is applied, the problems and loop holes of automated reference retrieval systems can be overcome. Eventually such systems will provide a realistic answer to libraries' rapidly expanding problem of both bibliographic control as well as accessibility of information.

BEST COPY AVAILABLE

APPENDIX A

BIOGRAPHICAL REFERENCE WORK DATA BASE

This document was processed for the ERIC Document Reproduction Service by the ERIC Clearinghouse at Stanford. We are aware that some pages probably will not be readable in microfiche or in Hardcopy form. However, this is the best available copy, and we feel that the document should not be withheld from interested readers on the basis of these unreadable pages alone.

NOTE: The following pages of Appendix A represents only about 60 of the 170 titles in the data base.

TYPE BOOKS.FIL

```
D    +        15001900APPLETONS CYCLOPEDIA OF AMERICAN BIOGRAPHY 1500-1900
  (POR) REF 920.07 AP54
D    +        15001945DICTIONARY OF AMERICAN BIOGRAPHY 1500-1945 REF 920.07
 D65A
D    +        15001940CONCISE DICTIONARY OF AMERICAN BIOGRAPHY REF 920.07
D56A
D    +        19031974WHOS WHO IN AMERICA 1903-1974 REF 920.07 W62A
RS   +        19271950WHOS WHO IN COLORED AMERICA 1927-1950 REF 920.07 W62
AC
D    +        16071968WHOS WHO IN AMERICA 1607-1968 REF 920.07 W62AW
C    +        19581974 WHOS WHO OF AMERICAN WOMEN 1958-1974 REF 920.07 W62
W
D    +        16501905DICTIONARY OF AMERICAN PORTRAITS 1650-1905 REF 920.0
73 C49E
DE   I        18001897 AMERICAS SUCESSFUL MEN OF AFFAIRS 1900-1897 REF 920
.073 H14A (N.Y. &U.S. BUSINESSMEN
D    +        18751926 AMERICAN BLUEBOOK OF BIOGRAPHY 1875-1926 REF 920.07
3 H43A
D    +ZY      18501915 AMERICAN STATESMEN AND PUBLIC OFFICIAL YEARBOOK 185
0-1915 REF 920.072 H43S
D    +CEFGIA19001973 INTERNATIONAL CELEBRITY REGISTER 1900-1973 REF 920.
073 IN8
X    +        19351967 ITALIAN AMERICAN WHO 1935-1967 WHO REF 920.073 IT 1
2    +        19711971 ASIAN WHOS IN AMERICA 1971-1971 REF 920.073 L79A
3    -        19661967 MERITS WHOS WHO AMONG AMERICAN HIGH SCHOOL STUDENTS
 1966-1967 REF 920.073 M545
D    +        17001965 NATIONAL CYCLOPEDIA OF AMERICAN BIOGRAPHY 1700-1965
 REF 920.073 N21
DEFGH+        19701971 NATIONAL REGISTER OF PROMINENT AMERICANS AND INTERN
ATIONAL NOTABLES 1970-1971 REF 920.073 N211N
4    -        19691969 NATIONAL STUDENT REGISTER 1969-1969 REF 920.073 N21
3 COLLEGE AND UNIVERSITY STUDENTS
D    +        17001974THE HALL OF FAME FOR GREAT AMERICANS AT NEW YORK U.
REF 920.073 N43H
4    +        19651974 OUTSTANDING YOUNG MEN OF AMERICA 1965-1974 REF 920.
073 OU8
D    +        17001940 AMERICAN BIOGRAPHIES 1700-1940 REF 920.073 P92
```

BEST COPY AVAILABLE

BEST COPY AVAILABLE


```
D    +      18001900 PROGRESSIVE AMERICANS 1900-1900 REF 920.073 P943
D    -      16421767 BIOGRAPHICAL SKETCHES OF GRADUATES OF HARVARD U. 16
42-1767 (PDR) REF 920.073 S111
D    +      18001900 THE 20TH CENTURY BIOGRAPHICAL DICTIONARY OF NOTABLE
 AMERICANS 1900-1900 REF 920.073 T919
D    //     19061973 REGISTER OF THE DEPT. OF STATE 1906-1973 REF 920.07
3 UN3
D    +      19491954 WHO KNOWS AND WHAT AMONG AUTHORITES, EXPERTS AND TH
E SPECIALLY INFORMED 1949-1954 REF 920.073 W62
X    +      18501970 WHOS WHO IN POLISH AMERICA 1850-1970 REF 920.073 W6
3
E    +      19421975 WHOS WHO IN THE EAST 1942-1975 REF 920.07 W62 (MIDD
LE ATLANTIC & NE STATES)
F    +      19501974 WHOS WHO IN THE SOUTH AND SOUTHWEST 1950-1974 REF 9
20.075 W62 (SOUTHERN AND SOUTHCENTRAL STATES)
F    /      18301836 THE MEN WHO MADE TEXAS FREE 1830-1836 REF 920.0764
D646M (SIGNERS OF TEXAS DEC. OF IND.)
F    +      19491968 WHOS WHO IN TEXAS TODAY 1968-1969 REF 920.0764 W62
F    AF     19261926 THE CREATIVE ARTS IN TEXAS 1926-1926 REF 920.0764 S
M57C
G    +      19491973 WHOS WHO IN THE MIDWEST 1949-1973 REF 920.077 W62 (
CENTRAL & MIDWESTERN STATES)
G    +      18001960 DICTIONARY OF WISCONSIN BIOGRAPHY 1800-1960 REF 920
.0775 W753D
GH   +      19681972 PERSONALITIES OF THE WEST AND MIDWEST 1968-1972 REF
 920.078 P432
I    +      19171947 WHOS WHO IN THE NORTHWEST 1917-1947 REF 920.078 W62
 (PDR)
H    +      19491973 WHOS WHO IN THE WEST 1949-1973 REF 920.078 W62W (PA
CIFIC COASTAL & WESTERN STATES)
G    +      19241925 FOXS WHOS WHOS AMONG SOUTH DAKOTANS 1924-1925 REF 9
20.0783 W62
H    +      19231923 WHOS WHO IN THE ROCKIES 1923-1923 REF 920.0789 L233
W (PDR)  (COLORADO)
H    +      19531956 WHOS WHO IN COLORADO 1956-1956 REF 920.0788 W62
H    +      19401942 CAPITOLS WHOS WHO IN DENVER AND COLORADO 1940-1942
REF 920.0789 W63C
D    ML     19491973 ALA MEMBERSHIP DIRECTORY 1949-1973 REF 920.2 AM35D
```

MH +$/ 18901909 PORTRAIT,GENEALOGICAL AND BIOGRAPHICAL RECORD OF ST
ATE OF UTAH 1890-1902 R 920.0792 N21P
LMH + 19321933 UTAHS DISTINGUISHED PERSONALITIES 1932-1933 R 920.0
792 S1 47A
LMH + 18301930 LDS BIOGRAPHICAL ENCYCLOPEDIA 1830-1930 R M270 J451
B (POR) (UTAH & MORMON FIGURES)
LMH + 18471910 PIONEERS & PROMINENT MEN OF UTAH 1847-1910 R M270 E
S77 (POR) (INCLUDES MORMON FIGURES)
H + 18501913 PRESS REFERENCE LIBRARY 1650-1913 R 920.0794 P926W
(POR) (INCLUDES CALIF.FIGURES)
H + 19551973 WHOS WHO IN CALIFORNIA 1955-1973 R 920.0794 W62 (PO
R)
H + 19321929 WHOS WHO IN LOS ANGELES COUNTY 1928-1929 R 920.0794
W62L (POR)
H IJK 19631963 WHOS WHO EXECUTIVES IN CALIFORNIA 1963-1963 R 920.0
794 W62E (POR)
H + 19131913 WHOS WHO IN THE PACIFIC SOUTHWEST 1913-1913 R 920.0
794 W62P (INCLUDES CALF.& ARIZONA)
J + 19551955 ALASKA WHOS HERE, WHATS DOING, WHOS DOING IT 1955-1
955 R 920.0798 J49
D ML 19321972 WHOS WHO IN LIBRARY SERVICE 1933-1972 R 920.2 W62
C CR 19741974 FOREMOST WOMEN IN COMMUNICATIONS 1974-1974 R 920.7
F761 (INCLUDES BROADCASTING, ADVERTISING, PUBLIC RELATIONS)
C + 16071950 NOTABLE AMERICAN WOMEN 1607-1950 R 920.7 N843
C +L43 19661973 OUTSTANDING YOUNG WOMEN OF AMERICA 1966-1973 R 920.
7068 O98 (INCLUDES EDUCATORS, NURSES, HOMEMAKERS)
C + 19691972 2000 WOMEN OF ACHIEVEMENT 1969-1972 R 920.7 T93 (PO
R)
C + 19001906 A WOMEN OF THE CENTURY; BIOG. SKETCHES OF AM. WOMEN
1800-1900 R 920.7 W65W (POR)
F + 19701973 PERSONALITIES OF THE SOUTH 1970-1973 R 920.975 P432
D 34=/ 19641966 BIOGRAPHICAL DICTIONARY OF PARA PSYCHOLOGY 1964-196
6 R 921.03 B52 (INCLUDES PSYCHOLOGIST & PHYSICIANS)
D CV 19421942 BIOGRAPHICAL ENCYCLOPEDIA OF PHILOSOPHY 1942-1942 R
921.03 W62
D =/ 19701973 AMERICAN PSYCHOLOGICAL ASSN. 1970-1973 R 921.1 AM35

BEST COPY AVAILABLE

BEST COPY AVAILABLE

APPENDIX B

COMPUTER PROGRAM - "LOOKUP"

The real core of the program is contained in lines 350 through 800. Lines prior to 350 serve to establish variables and set formats. For example, the data base "books.fil" is established, as are the word and letter variables, "end-word," "search-key," "found-it," "place-key," "occupation-key," "date," "I," and "J." The program allows the "place field" (aforementioned as miscellaneous field) to occupy five spaces, the "occupation field" to occupy seven spaces, and the "date field" to occupy eight spaces, (four spaces for the beginning date and four spaces for the ending date). The program also allows the title and miscellaneous information to occupy 95 spaces.

The program is written so that the computer prints or displays on the screen certain phrases that are helpful to the execution of the program. For example, the computer displays, "Type in search-key in the following format, Where P=Place, O=Occupation, DDDD=Year." This helps the user to know in what order he is to type in his search profile.

Lines 370 through 780 are composed of a series of simple loops which compare each symbol in the patron profile or search-key with each symbol in the data base. For example, suppose the following search-key were typed into the computer for a search:

M+1900

(Incidentally, this means a biography of a Mormon of any occupation who lived in 1900.) Only two symbols and one year can be submitted for search at one time. The user could not, for example, type in M+L1900, but he could make as many repeated searches, one after the other, as he desired, using both "specific" as well as general symbols. The computer will take the first symbol of the search-key, which in this case is the letter M, and compare that letter with all of the symbols located in the first five

BEST COPY AVAILABLE

spaces of the miscellaneous fixed field of each of the 170 works. The computer will determine if the letter M matches with any other M. If there is a match, those works are tagged and set aside.

The machine then takes the next symbol of the search-key, which in this case is + and compares that symbol with the seven symbols or spaces in the occupation fixed field, which immediately follows the five spaces of the miscellaneous field. However, the computer will only attempt to match the + with those works that were matched with the M. In this way, the 170 works are narrowed down each time a match is made with each symbol. If there is a match with the symbol +, the computer then tags or identifies those works and sets them aside.

The date within the search-key is taken next. In this case the date is 1900. The computer attempts to match this date with those works that have been set aside and have matched both with the M and the +. The machine takes the two dates, within the date field, and simply says, "Is the search-key date (which is 1900) more than the beginning date and less than the ending date?" If so, then a match has taken place.

Those works that have a match on all three fields, are tagged and set aside in a variable called "title-table." The program then instructs the computer to print out "title-table," which is a list of those works, taken from the 170 works, that matched on all three areas of the search-key. Any other work that matched on two areas or less is not retrieved.

It was mentioned earlier that one apparent disadvantage of the indexing was that a substantial amount of works were indexed in terms of only one year, such as 1923 to 1923. If the patron does not type in the precise date within his search-key then a match will not occur. Thus, many works could not be retrieved simply because the year within the

search-key did not happen to match the years indexed in the reference works. This is certainly a disadvantage; however, the program is written so that a user may or may not make a search with a date. At the beginning of the program the computer is instructed to type out, "search for date?" If the user types in Y, then a date must be entered in the search-key; however, if the user types in N, then a date must not be entered in the search-key. This is a significant advantage in the retrieval rule.

Initially, a user should type in a date. Moreover, he should also be as specific as possible within the classification scheme and the descriptors. If there is a "no match," or if there are only a few matches, then the user should make another run and should be more general in his search-key, and if desired, may search without a date. This will increase the number of titles retrieved. Even though more titles are retrieved without a date, it is easy for the user to separate the relevant from the irrelevant since the dates the work covers are displayed with the title.

The following example will help to illustrate what happens when the user searches specifically, then, more generally, and finally without a date. The search-key D41900 was typed into the computer, which means a biography of a United States physican living in 1900. No references were found with that search-key so it was changed and resubmitted as D41920. Still no match occured so the user omitted the date altogether, and entered D4. The following works were retrieved using that search-key:

Biographical Directory of Para-psychologists.
Who's Who in Medicine.
Biographical Dictionary of Eminent American Physicians.

Still desiring more works, the user altered his search-key to D3, which means a biography of United States men of medicine. This made the search-

BEST COPY AVAILABLE

key even more general. The following works were retrieved using the more general search-key:

Biographical Directory of Para-psychologists.
Who's Who in Medicine
Biographical Dictionary of Eminent American Physicians.
Outstanding Young Women of America (includes Nurses)
Who's Important in Medicine
American Medical Biography
Biographies in the New York Academy of Medicine

The ability of the user to be both specific and general, as well as to search both with or without a date, is extremely valuable.

A final aspect of the program needs explanation. How will the user know if there is no match on the search-key? The program instructs the machine to print out the phrase, "Unable to locate any references with that key; if you desire another search, type "MORE," otherwise type "END." If the patron types "END" the program will not run again; however, if he types "MORE," the program will repeat itself. This allows the patron to execute, very quickly and as many times as desired, the search routine. Depending upon the complexity of the search, the user usually will only spend a matter of seconds on the terminal to arrive at a satisfactory answer.

BEST COPY AVAILABLE


```
10 IDENTIFICATION DIVISION.
20 PROGRAM-ID. LOOKUP.
30 ENVIRONMENT DIVISION.
40 INPUT-OUTPUT SECTION.
50 FILE-CONTROL.
60          SELECT LIBRARY-FILE ASSIGN TO DSK
65                  RECORDING MODE ASCII.
70 DATA DIVISION.
80 FILE SECTION.
90 FD     LIBRARY-FILE
100        VALUE OF ID "BOOKS FIL".
110 01     BOOK.
120        05     PLACE OCCURS 5 TIMES PIC X.
130        05     OCCUPATION OCCURS 7 TIMES PIC X.
140        05     BEGINNING-DATE PIC 9(4).
150        05     ENDING-DATE PIC 9(4).
160        05     TITLE-ETC PIC X(95).
170 WORKING-STORAGE SECTION.
180 77      FOUND-IT          PIC 9 VALUE 0.
190 77      I                 PIC 99 VALUE 0.
200 77      J                 PIC 99 VALUE 0.
210 77      END-WORD          PIC X(4) VALUE SPACES.
215 77      YORN              PIC X.
220 01      SEARCH-KEY.
230         05     PLACE-KEY       PIC X.
240         05     OCCUPATION-KEY  PIC X.
250         05     DATE            PIC 9(4).
260 01      TABLE.
270         05     TITLE-TABLE OCCURS 20 TIMES PIC X(95).
290 PROCEDURE DIVISION.
290 OPEN-FILE.
295         OPEN INPUT LIBRARY-FILE.
300 DISPLAY-HEADS.
305       DISPLAY "SEARCH FOR DATE?".
308       ACCEPT YORN.
310        DISPLAY "TYPE IN SEARCH KEY IN THE FOLLOWING FORMAT:".
320        DISPLAY " ".
```

```
330         DISPLAY "                    PODDDD".
340         DISPLAY "WHERE P=PLACE, O=OCCUPATION, DDDD=YEAR,"
350 ACCEPT-SEARCH-KEY.
360         ACCEPT SEARCH-KEY.
370 SEARCH-LIBRARY-FILE.
380         READ LIBRARY-FILE
390                 AT END GO TO ANY-MORE.
400         MOVE 0 TO FOUND-IT.
410         PERFORM SEARCH-PLACE
420                 VARYING I FROM 1 BY 1
430         UNTIL I > 5 OR FOUND-IT = 1.
440         IF FOUND-IT NOT EQUAL 1
450                 GO TO SEARCH-LIBRARY-FILE.

460         PERFORM SEARCH-OCCUPATION
470                 VARYING I FROM 1 BY 1
480                 UNTIL I > 7 OR FOUND-IT = 2.

490         IF FOUND-IT NOT = 2
500                 GO TO SEARCH-LIBRARY-FILE.
505         IF YORN = "Y" NEXT SENTENCE
506         ELSE GO TO MOVE-TITLE.
510         PERFORM DATE-COMPARE.



520         IF FOUND-IT NOT=3
530                 GO TO SEARCH-LIBRARY-FILE.
535 MOVE-TITLE.
540         ADD 1 TO J.
550         MOVE TITLE-ETC TO TITLE-TABLE (J).
560         GO TO SEARCH-LIBRARY-FILE.
```

BEST COPY AVAILABLE

```
570 ANY-MORE.
580         PERFORM WRITE-TABLE
590                 VARYING I FROM 1 BY 1
600                 UNTIL I > J.
605         IF J = 0
604                 DISPLAY "UNABLE TO LOCATE ANY REFERENCES WITH THAT KEY.".
610         MOVE 0 TO J.
620         DISPLAY "IF YOU DESIRE ANOTHER SEARCH, TYPE MORE".
630         DISPLAY "OTHERWISE TYPE END."
640         ACCEPT END-WORD.
645         CLOSE LIBRARY-FILE.
650         IF END-WORD = "MORE"
660                 GO TO OPEN-FILE.
680         STOP RUN.
690 SEARCH-PLACE.

700         IF PLACE (I) = PLACE-KEY
710                 ADD 1 TO FOUND-IT.
720 SEARCH-OCCUPATION.
730         IF OCCUPATION (I) = OCCUPATION-KEY
740                 ADD 1 TO FOUND-IT.
750 DATE-COMPARE.
760         IF DATE NOT < BEGINNING-DATE
770                 IF DATE NOT > ENDING-DATE
780                         ADD 1 TO FOUND-IT.
790 WRITE-TABLE.
800         DISPLAY TITLE-TABLE (I).
```

BEST COPY AVAILABLE

APPENDIX C

SEARCH EXAMPLES

Chart no. 2

Search Examples

Search-key:

D91970 (U. S. Chemist in 1970)

Works Retrieved:

American Men of Science 1906-1971 R 925 Am35 (Social, Behavioral, Physical and Biological Sciences)

Search-key:

D9 (U. S. Chemist)

Works Retrieved:

American Men of Science 1906-1971 R 924 Am35 (Social, Behavioral, Physical, and Biological Sciences)

Chemical Who's Who 1928-1955 R 925.4 C42

Search-key:

D61970 (U. S. Physical and Biological Scientists in 1970)

Works Retrieved:

American Men of Science 1906-1971 R 925 Am35 (Social, Behavioral, Physical and Biological Sciences)

Dictionary of Scientific Biography 1600-1970 R 925 D361

American Entomologists 1750-1971 R 925.9 M296a (por)

Leaders of American Conservation 1900-1971 R 926.349 O5941

Search-key:

X+1965 (European-Americans of all occupations in 1965)

Works Retrieved:

Italian American Who's Who 1935-1967 R 920.073 It1

Who's Who in Polish America 1850-1970 R 920.073 W62

Chart no. 2

Search Examples

Search-key:

3- (U. S. Students 0 to 18 years old)

Works Retrieved:

Merits Who's Who Among American High School Students 1966-1967 R 920.073 M545

Search-key:

C+1966 (U. S. Women of all occupations in 1966)

Works Retrieved:

Who's Who of American Women 1956-1974 R 920.07 W62w

Outstanding Young Women of America 1966-1973 R 920.7062 Ou6

Search-key:

M+1898 (Mormons of all occupations in 1898)

Works Retrieved:

Portrait, Genealogical and Biographical Record of the State of Utah 1890-1902 R 920.0732 N21p

LDS Biographical Encyclopedia 1830-1930 R M270 J451b (por)

Pioneers and Prominent Men of Utah 1847-1910 R M270 B451b

Search-key:

DF1970 (U. S. Authors in 1970)

Works Retrieved:

International Celebrity Register 1900-1973 R 920.073 In8

Junior Book of Authors 1900-1972 R 929.08 K96j (por)

Authors and Writers Who's Who of English Speaking World 1934-1971 R 928.2 Au91

Chart no. 2

Search Examples

Search-key:

DF (U. S. Authors)

Works Retrieved:

International Celebrity Register 1900-1973 R 920.073 In8

A Biographical Companion to the Arts 1700-1965 R 927 At63

Dictionary of American Authors 1700-1905 R 928 Ad19

The American Literary Yearbook 1850-1919 R 928 Am35

The Biographical Dictionary of Contempory Poets 1850-1938 R 928 B52

American Authors 1600-1900 R 928 K96am

Authors of Books for Young People 1850-1967 R 928 W215a

Who's Who Among North American Authors 1921-1939 R 928 W62n

Allibones Dictionary of Authors 1600-1891 R 928.08 A158

Contemporary Authors 1900-1962 R 928.08 0767

Junior Book of Authors 1900-1972 R 929.08 K96t (por)

20th Century Writing 1900-1969 R 928.08 R394t

Authors and Writers Who's Who of the English Speaking World 1934-1971 R 928.2 Au61

BEST COPY AVAILABLE

Chart no. 2

Search Examples

Search-key:

H+ (Western United States persons of all occupations)

Works Retrieved:

National Register of Prominent Americans and International Notables 1970-1971 R 920.075 N811

Personalities of the West and Midwest 1962-1972 R 920.079 P432

Who's Who in the West 1949-1973 R 920.078 W62w

Who's Who in the Rockies 1926-1933 R 920.0783 L233w (por)

Who's Who in Colorado 1958-1968 R 920.0788 W62

Capitol's Who's Who in Denver and Colorado 1940-1942 R 920.0788 W62c

Who's Who in New Mexico 1937-1937 R 920.0739 W63

Who's Who on the Pacific Coast 1913-1949 R 920.079 W62

Portrait. Genealogical and Biographical Record of the State of Utah. 1890-1902 R 920.0792 N21f

Utah's Distinguished Personalities 1932-1933 R 920.0792 Si47a

LDS Biographical Encyclopedia 1930-1950 R M270 J451b (por)

Pioneers and Prominent Men of Utah 1847-1910 R M270 E877 (por)

Press Reference Library 1850-1913 R 920.0794 P926w (por) (includes California figures)

Who's Who in California 1955-1973 R 920.0794 W62 (por)

Who's Who in Los Angeles County 1928-1929 R 920.0794 W621 (por)

Who's Who in the Pacific Southwest 1912-1913 R 920.0794 W52p

BEST COPY AVAILABLE

SOURCES CONSULTED

Books Consulted

Maron, M.E. An Information Processing Laboratory for Education and Research in Library Science, Phase I. Berkeley: Institute of Library Research, University of California, Berkeley, 1969.

Periodicals Consulted

Chen, Ching-chich. "Subject Reference Lists Produced by Computer." Journal of Library Automation 1 (September 1969): 178-97.

Finzi, J.C. "Computer-based Reference Services; A Pre-conference Institute of the Reference Services Division of ALA." Library of Congress Information Bulletin 30 (July 8, 1971): 111-14.

Geller, E. "Hypothetic Dialogues: The RSD Preconference." Library Journal 96 (August 1971): 2450-51.

Henn, S. "Timely Tips and Automated Trends in Reference Service." Virginia Librarian 16 (Summer 1969): 20-1.

Jones, J.C. "Impacts on Reference: Census and the Computer." RQ 12 (Spring 1973): 247-50.

Meredith, Joseph C. "Machine Assisted Approach to General Reference Materials." Journal of the American Society for Information Science 22 (May-June 1971): 176-86.

Mignon, Edmond. "Information Science in the Teaching of Traditional Reference Service." Proceedings of the American Society for Information Science 8 (1971): 143-46.

Prodrick, R.G. "Automation Can Transform Reference Services." Ontario Library Review 51 (September 1967): 145-50.

Schultz, Claire K. "Automation of Reference Work." Library Trends 12 (January 1964): 413-24.

Shera, Jesse. "Automation and the Reference Librarian." RQ 3 (July 1964): 3-4.

Weil, Cherie B. "Automation Retrieval of Biographical Reference Books." Journal of Library Automation 1 (December 1968): 239-49.